

# Assignment 10.2 on Logistic Regression Model (Thoracic Surgery and Binary Classifier Data)

Basak Atanu

05-17-2022

```
library(foreign)

setwd("C:\\Users\\atanu\\Documents\\BellevueUniversity_MSDS\\DSC520\\Repository\\dsc520_")
thoraric <- read.arff("data\\ThoraricSurgery.arff")
#head(thoraric)
```

lets fit the logistic regression model.

```
model <- glm(Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + + PRE8 + PRE9 + PRE10 + PRE11 +
              PRE14 + PRE17 + PRE19 + PRE25 + PRE30 + PRE32 + AGE, data=thoraric, family='binomial')
summary(model)
```

```
##
## Call:
## glm(formula = Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + +PRE8 +
##     PRE9 + PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 +
##     PRE32 + AGE, family = "binomial", data = thoraric)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6084  -0.5439  -0.4199  -0.2762   2.4929
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.655e+01  2.400e+03  -0.007  0.99450
## DGNDGN2      1.474e+01  2.400e+03   0.006  0.99510
## DGNDGN3      1.418e+01  2.400e+03   0.006  0.99528
## DGNDGN4      1.461e+01  2.400e+03   0.006  0.99514
## DGNDGN5      1.638e+01  2.400e+03   0.007  0.99455
## DGNDGN6      4.089e-01  2.673e+03   0.000  0.99988
## DGNDGN8      1.803e+01  2.400e+03   0.008  0.99400
## PRE4         -2.272e-01  1.849e-01  -1.229  0.21909
## PRE5         -3.030e-02  1.786e-02  -1.697  0.08971
## PRE6PRZ1     -4.427e-01  5.199e-01  -0.852  0.39448
## PRE6PRZ2     -2.937e-01  7.907e-01  -0.371  0.71030
## PRE7T         7.153e-01  5.556e-01   1.288  0.19788
## PRE8T         1.743e-01  3.892e-01   0.448  0.65419
```

```
## PRE9T      1.368e+00  4.868e-01  2.811  0.00494 **
## PRE10T     5.770e-01  4.826e-01  1.196  0.23185
## PRE11T     5.162e-01  3.965e-01  1.302  0.19295
## PRE140C12  4.394e-01  3.301e-01  1.331  0.18318
## PRE140C13  1.179e+00  6.165e-01  1.913  0.05580 .
## PRE140C14  1.653e+00  6.094e-01  2.713  0.00668 **
## PRE17T     9.266e-01  4.445e-01  2.085  0.03709 *
## PRE19T    -1.466e+01  1.654e+03 -0.009  0.99293
## PRE25T    -9.789e-02  1.003e+00 -0.098  0.92227
## PRE30T     1.084e+00  4.990e-01  2.172  0.02984 *
## PRE32T    -1.398e+01  1.645e+03 -0.008  0.99322
## AGE       -9.506e-03  1.810e-02 -0.525  0.59944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 395.61  on 469  degrees of freedom
## Residual deviance: 341.19  on 445  degrees of freedom
## AIC: 391.19
##
## Number of Fisher Scoring iterations: 15
```

according to the summary, PRE9 (Dyspnoea before surgery), PRE14: T in clinical TNM - size of the original tumour with OC14 which means large size of tumour have greater affect on servical rate.

```
res <- predict(model, thoraric, type='response')
confmatrix <- table(actual_value=thoraric$Risk1Yr, predicted_value = res>0.5)
```

## Accuracy

```
(confmatrix[[1,1]] + confmatrix[[2,2]]) / sum(confmatrix)
```

```
## [1] 0.8361702
```

## Reading binary classifier data.

```
setwd("C:\\Users\\atanu\\Documents\\BellevueUniversity_MSDS\\DSC520\\Repository\\dsc520_")
classifier_data <- read.csv("data\\binary-classifier-data.csv")
```

## Fitting logistic regression model.

```
model <- glm(label ~ x + y, data=classifier_data, family='binomial')
summary(model)
```

```
##
## Call:
## glm(formula = label ~ x + y, family = "binomial", data = classifier_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3728  -1.1697  -0.9575   1.1646   1.3989
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.424809   0.117224   3.624  0.00029 ***
## x           -0.002571   0.001823  -1.411  0.15836
## y           -0.007956   0.001869  -4.257  2.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2075.8  on 1497  degrees of freedom
## Residual deviance: 2052.1  on 1495  degrees of freedom
## AIC: 2058.1
##
## Number of Fisher Scoring iterations: 4
```

Accuracy of logistic regression classifier.

```
res <- predict(model, classifier_data, type='response')
confmatrix <- table(actual_value=classifier_data$label, predicted_value = res>0.5)
(confmatrix[[1,1]] + confmatrix[[2,2]]) / sum(confmatrix)
```

```
## [1] 0.5834446
```